

Connecting to Smart Cities: Analyzing Energy Times Series to Visualize Monthly Electricity Peak Load in Residential Buildings

Abstract. Rapidly growing energy consumption rate is considered an alarming threat to economic stability and environmental sustainability. There is an urgent need of proposing novel solutions to mitigate the drastic impact of increased energy demand in urban cities to improve energy efficiency in smart buildings. It is commonly agreed that exploring, analyzing and visualizing energy consumption patterns in residential buildings can help to estimate their energy demands. Moreover, visualizing energy consumption patterns of residential buildings can also help to diagnose if there is any unpredictable increase in energy demand at a certain time period. However, visualizing and inferring energy consumption patterns from typical line graphs, bar charts, scatter plots is obsolete, less informative and do not provide deep and significant insight of the daily domestic demand of energy utilization. Moreover, these methods become less significant when high temporal resolution is required. In this research work, advanced data exploratory and data analytics techniques are applied on energy time series. Data exploration results are presented in the form of heatmap. Heatmap provides a significant insight of energy utilization behavior during different times of the day. Heatmap results are articulated from three analytical perspectives; descriptive analysis, diagnostic analysis and contextual analysis.

Keywords: Energy Efficiency, Smart Buildings, Data Analytics, Heatmap.

1 Introduction

In recent years, energy data analytics has got tremendous attention of researchers, economists, industrialists, and policy makers from all over the world. This could be because of the shortage of natural resources, environmental destruction, or proliferation of energy demand due to the development of urban cities. Confronted, with this rapid increase of energy demand, the researchers and scientists are finding greater interest to design and develop advanced techniques and methods that can help us to cope with energy crises or at least to mitigate its worst consequences.

Moreover, the rapidly increasing energy consumption rate poses an alarming threat to the worldwide environmental sustainability and economic stability. International Energy Agency's (IEA) statistics reveal that 32% of the total final energy is being consumed by the buildings [1]. This percentage is even higher in non-industrial areas. The fact that how people consumes energy depends on human behaviour and other social, economic, environmental and geographical factors [2].

In recent years, energy efficiency and saving strategies have become a priority objective for energy policies due to the proliferation of energy consumption and CO₂ emission in the built environment. According to statistics 40% of all primary energy

is being consumed in and by the buildings [3]. International Energy Agency (IEA) in [5] claims that “*Energy efficiency is a critical tool to relieve pressure on energy supply and it can also mitigate in part the competitive impacts of price disparities between regions*”.

Analyzing energy patterns and identifying variations in energy usage with the help of data mining techniques will help to build energy efficient buildings. It is evident in the past 40 years that increasing energy efficiency of the buildings helps not only to combat the climate changes but also to reduce the energy consumption [4].

Furthermore, this research work presents a framework that brings multi-domain knowledge to an interdisciplinary project to solve the unaddressed or partially addressed issue in the domain of energy efficient smart buildings. In doing so, this research work elucidates the importance of mapping multi-domain experts’ opinion to develop the new policies in deploying the significant changes. This new approach that combines social, economic, behavioural and psychological, environmental, statistical and computational phenomena offers a dynamic and compelling framework for designing energy efficient buildings. This research work also acts as a bridge to fill the communication gap between research community and the policy makers to make intelligent decisions based on scientific evidence.

1.1 Times Series Analysis

In time series analysis concern lies in forecasting a specific quantity given that the variations in that quantity over time are already known. While, other predictive models that do not involve time series mainly focus on analysing a cross-sectional area of the data which do not have time variance component. As stated by Hilda *et al.*, in [6], “*When a variable is measured sequentially in time over or at a fixed interval (sampling interval) the resulting data represents a time series*”. They further elaborated that time series is a collection of observations arranged in a natural order where each observation is associated with a particular instance or interval of time.

More specifically, time series, compared to common data, holds natural temporal ordering where common data does not necessarily have natural ordering of the observations. Furthermore, Millan *et al.*, [7] defined time series analysis as a process of using statistical techniques to model and explain a time-dependent series of data points. Whereas, time series forecasting uses a prediction model to forecast the future events based on the past events.

This research work also presents the application of different kinds of analytical and visualization techniques to understand energy utilization patterns in residential building. Data analytical results are visualized in the form of heatmap. Heatmap results are articulated from three different analytical perspectives as descriptive analysis, diagnostic analysis and contextual analysis. Rest of the paper is structured as; state of the art work is presented in section 2 followed by methodological framework in section 3. Exploratory data analytical techniques are elaborated in section 4; whereas section 5 details the data that is used in this research work along with data preprocessing techniques. Application of heatmap examples are explained in section 6. Section 7 provides brief summary of the work along with conclusion and future research work.

2 Literature Survey

Platchkov and Pollitt in their paper [8] critically analysed and overviewed the longer run trends of increasing global electricity demands and explain the potential impact in the UK electrification. They claimed that the underlying resources cost for the energy that is being used in different times of the day or the year changes accordingly. For instance, on an off-peak day the price per megawatt hour (MWh) in the power market does not rise above £50/MWh, however, on the peak day the price may reach to £800 for half hour periods across a 24-hour period. This implies that, for median days there is a comparatively great incentive of using electricity during night time. The main emphasis of their research work is that the demand will increase steadily over time but the possible coping solution is to shift the energy demand to off-peak time.

Therefore, a small demand response, either by reducing the consumption or by shifting it to the cheaper time can make a significant difference in cost for residential as well as for commercial buildings. This shows the significance of shifting demand to off-peak time which is also called load balancing. Furthermore, figuring out the factors that trigger the peak energy demand for a specific period of time in a building could potentially help to improve building's heating, ventilation and air conditioning (HVAC) system. Together with this, sudden peak in energy consumption can be because of some mal-functioning or some unexceptional human behavior. Finding possible causes of high energy demand for a certain period of time can possibly lead to find appropriate solutions for it and ultimately a control in energy demand. Understanding this demand and supply behavior in residential areas will further support the sustainable and renewable energy technology.

David in his paper [2] states that selecting key variables and interactions is therefore an important step in achieving more accurate predictions, better interpretations, and identifications of key subgroups in the energy datasets for further analysis. Jenkins *et al.*, [8] visualize energy data to examine the monthly demand of substations and synthesized equivalent. Walker and Pokoski [9] developed a model of residential electric load where they introduced the psychological factors based on a person's availability that can affect the individual use of electrical appliances at a given time. Before that, in early nineties, Capasoo *et al.*, [10] applied bottom up approach to develop "Capasoo Model". This model uses the socioeconomic and demographic data, for instance, the stock of appliances and their usage pattern in a household to model a load curve. This load shape shows the relationship between the demand of residential customers and the psychological and behavioral factor of the house occupants. Later in 2002, [11] Willis used the bottom up approach to model the typical demand forecasting scheme for the individual customers.

3 Methodological Framework

The proposed methodological framework, for energy efficient smart buildings, provides foundation for complex, diverse, contextually aware, eco-driven and intelligently monitored nature of energy demand that frequently requires a multi domain, interdisciplinary approach into research. This framework articulates the

energy efficiency paradigm with respect to four significant attributes that should be considered to improve end-use energy efficiency and to reduce energy demand. The embedded features are predicated on the issues related to global climate change, social behavior, economic productivity, and modelling the exceptionally large energy datasets to explore and interpret the interesting, useful patterns of energy usage.

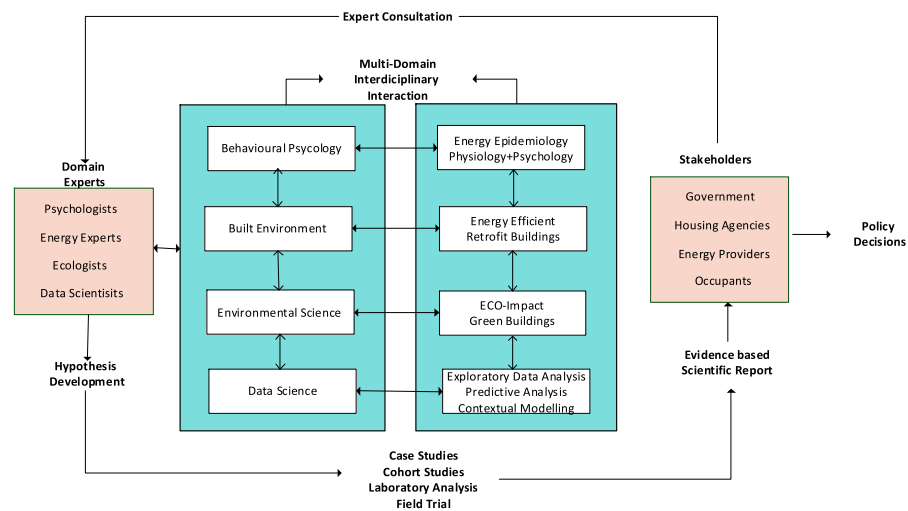


Fig. 1. A methodological framework for cross disciplinary knowledge exchange to exploit the design and development of energy efficient smart buildings.

The first crucial step to achieve a particular milestone is to identify and analyze the problems, issues and concerns of different stakeholders in order to develop a shared vision with common understanding and clear targets. The most important factor that should be considered in constructing the smart buildings or smart cities is “human beings”, which means, everything that we construct should be human oriented. Creating a comprehensive roadmap will help us to focus on high-return predictive analytics with clear pre-defined destinations and achievable milestones which is a starting point for gaining a better understanding of customer’s requirements.

Hence, as a part of this research work, one of the milestones is to classify the prerequisites to provide a foundation to develop a globally acceptable socio-technical strategy for building the smart buildings and smart cities. This will help to tackle all the issues that are in mutual interests of different stakeholders. Since, this is a long term ongoing project, this first part of the research work has already been accomplished and published [12].

Our next research question is what is the role of data science in the design and development of energy efficient smart buildings. In this research work, advanced analytical methods and visualization techniques are used to explore complex energy datasets in order to understand energy consumption patterns of a residential building.

4 Data Exploration- A Possible Solution

Data could be explored, analyzed, visualized and described at different level of maturity. Most of the existing literature reveals four (4) informative levels of data exploration depending on the complexity of the case studies under question. These are recognized as descriptive analysis, diagnostic analysis, predictive analysis and prescriptive analysis [1].

However, what is mostly neglected in most of the case studies analysis is to understand the circumstances in which a particular thing has happened. This is usually called contextual awareness. Credibility of the results could only be attained by linking the outcome of a particular analysis with certain situation in which it occurs. We are recommending contextual analysis as complementary method to describe any analytical results. Therefore, data analytical types could be described from five different perspectives as listed in **Table 1**.

Table 1. Data Exploration Types, Description and Examples.

Analytic Type	Description	Example
Descriptive Analysis	What is happening?	Historical data reports
Diagnostic Analysis	Why did it happen?	Fault Detection
Predictive Analytics	What is likely to happen?	Cost Prediction
Prescriptive Analysis	What should we do about it	Cost Optimization
Context Analysis	In which circumstances this happened?	Situation dependency

As mentioned earlier, this research work aims to understand energy utilization patterns in a residential building to identify any unusual data behavior and their reasons. Hence, the analysis will be carried out from three different perspectives as;

- Understanding energy utilization patterns → Descriptive Analysis
- Identifying extreme or abnormal data values → Diagnostic Analysis
- Finding the root cause of normal and extreme behavior → Context Analysis

5 Data Description

For this preliminary research, data is collected for 32 different houses in the area of Manchester in different domains. In the domain of Building Information data is collected for Archetype of the buildings, their Age, Addresses as longitude and latitude, Class, Construction type, Ownership of the buildings, Floor area and Air test. Fifteen various kinds of archetypes of the buildings were found in that area named as BISF, Brick and block, Detached 1980s brick and block, End terrace pre1919 solid wall, Flat Wimpey-no-finess non-trad, Mid terrace pre 1919 solid wall, Semi-detached pre 1919 solid wall, Semi-detached 1919 solid wall, Semi-detached 1920s solid wall, Semi-detached 1930s solid wall, Semi-detached 1970s brick and block cavity, Semi-detached pre 1800 brick, Terraced pre 1919 solid wall and Wates. Age of the building is categorised as 1920s, 1930s, 1950s, 1960s, 1970s, 1980s, pre 1800, pre 1919. Classes are defined as Detached, End-terraced, Flats, Mid-Terraced, Semi-detached.

Construction type is recognized as Traditional and Non-traditional. Floor area is measured in square meters (m^2) which is further classified into three sections as Small ($<50 m^2$), Medium ($50-100 m^2$) and Large ($>100 m^2$). Air permeability results for air leakage test are categorised into three sections as ($<5 m^3/(m^2.h)$), ($5-10 m^3/(m^2.h)$), ($>10 m^3/(m^2.h)$).

Demographic Information that is collected in the domain of Human Information constitutes their Age, Gender, Family Composition and their Health Status. Family composition is further recognised as Single occupants, Working couples, Small family, Small family of three, Family of four, Family of five, Family of six, Retired singles, Retired couples, Family of five with retired couples, and short term occupants with complex needs. In the Services domain data is collected for electricity and gas usage in KWH/m^2 for one complete year. Electricity data is clustered into three sections as ($<35 KWH/m^2$), ($35-40 KWH/m^2$), and ($>40 KWH/m^2$) whereas, gas data is also clustered into three sections as ($<120 KWH/m^2$), ($120-140 KWH/m^2$), and ($>140 KWH/m^2$).

5.1 Data Preprocessing

To understand data distribution, to find any outliers due to some extreme external behavior or malfunction in the sensor devices and to prepare data for analyzing and visualizing heatmap, energy dataset is preprocessed. At first, Cumulative Distribution function (CDF) is applied on datasets to understand the probability of random variables in the datasets. Equation (1) and (2) represents the cumulative distribution function $F(n)$ which is an estimate of the true CDF. It is found by making no assumptions about the underlying distribution.

$$F(t) = P(X \leq t) \quad (1)$$

$$F_n(t) = \frac{\# \text{ of Sample Values } \leq t}{n} \quad (2)$$

Fig. 2 a) is the visual representation of CDF for temperature dataset for whole building over one month. This includes hallway, Lounge and bedrooms. However, **Fig. 2 b)** represents boxplot diagram to understand the extreme data behavior which is sometime because of some malfunction in the devices.

Temperature dataset is collected for complete one year for all 40 buildings. However, to keep the analysis and visualization simple for this research work a dataset of one month (January) is selected for one residential building. Dataset is prepared by applying some functions from R¹ packages such as *lubridate*, *timeseries*, and R classes *POSIXct* and *POSIXlt*.

After discussions, it is decided to resample the datasets for different timestamp to remove any suspicious or null value. Temperature dataset was collected after each five seconds at first for 24 hours in a day for one year. However, to reduce the probability of any outliers, dataset was converted to each half an hour. This removed the probability of any extreme/malfunction data behavior that could affect the results. After that, heatmap algorithms are designed using R package *ggplot2*. Detail about heatmap application is articulated in the next section.

¹ <https://www.r-project.org/>

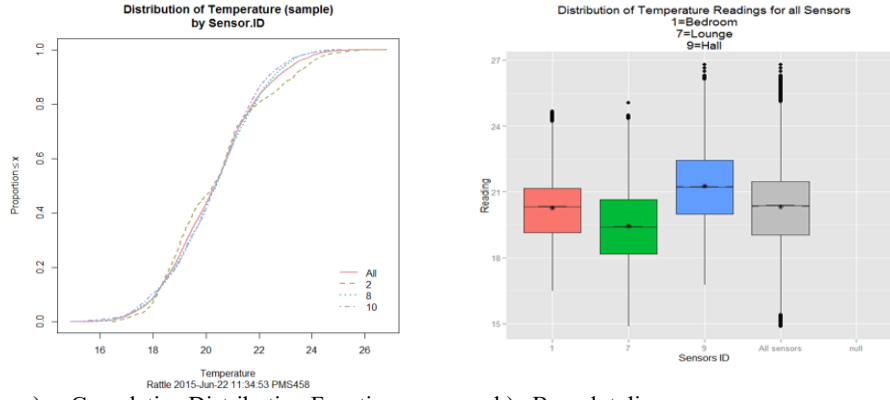


Fig. 2. a) Cumulative distribution of dataset. b) Outliers identification with Boxplot diagram.

6 Peak Identification- Heatmap example

Once data is preprocessed and cleaned, the next step is to visualize energy utilization patterns of a residential building. For this, we selected a building where a working couple was living. The idea is to understand the usual behavior of energy utilization for each day of a month. Also, apart from identifying their energy exploitation behavior, the intention was to diagnose if there are any extreme or unusual data patterns that could also be identified in the datasets.

As explained earlier, R library *ggplot2* is selected to design heatmap algorithm. **Fig. 3** provides visual representation of heatmap data values which are categorized from 0-2000 KWH and the color bar selected with dark blue, red and yellow colors where dark blue represents least data value and yellow represents extreme data value. Each data point in the heatmap presents a data value for half an hour which extends from 0-24 hours. However, y-axis represents each day of the month. Heatmap will help us to perform descriptive, diagnostic as well as contextual analysis.

As we can visualise in **Fig. 3**, there are some regular and some irregular energy utilisation patterns for each day in the whole month. As we can see in the figure, from 11 PM to 7:00 AM the data values range comes within blue band which identifies low energy usage at that time which is highlighted as night time in the figure. Then from 7:00 AM to around 11:30 AM there is comparatively higher usage of electricity which is probably due to the fact that everyone in the home is using electricity for normal house hold activities at that time of the day. This can be visualised as red colour squares in the figure. Then during the day time, again there is not much activity at home as compare to the night time. This probably because they have left the house for work. Then, between time span 5:30 PM to 11PM higher energy consumption could be visualised when usually everybody is at home and is engaged with different activities at home.

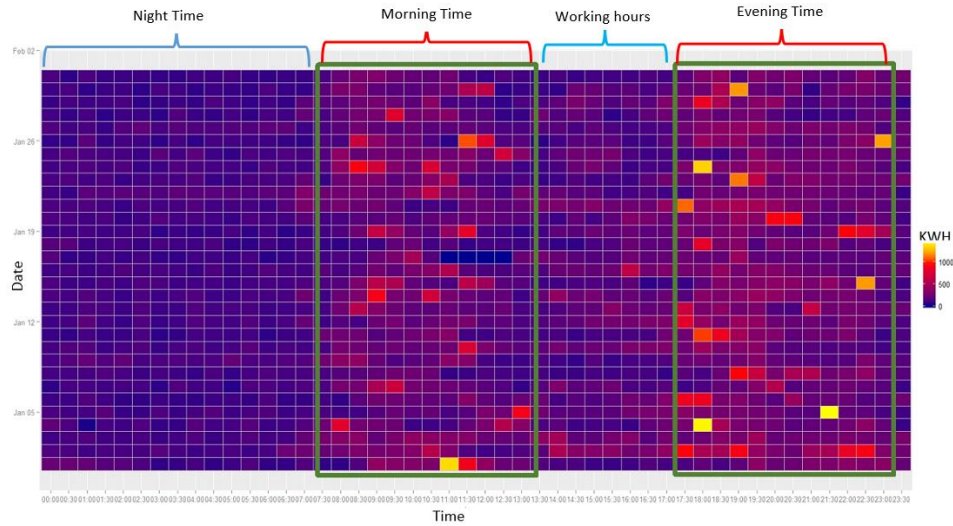


Fig. 3. Heatmap example to diagnose regular and extreme data behavior for a residential building.

Moreover, this is also evident from the description above that by linking the description of analytical results with its particular context actually helps to understand the reasons of least and higher electricity consumption at particular time of the day.

Apart from a normal energy utilisation patterns, some extreme data behaviour could also be visualised in the heatmap. For instance, all yellow points in the map tell us some extreme or abnormal energy utilisation behaviour. This implies that there could be some abnormality in the devices integrated in the house or this could be because of some unusual behaviour of the residents. Identifying abnormal or extreme behaviour in energy consumption patterns is called diagnostic analysis of the data. This also implies that further investigation could be recommended to find the root cause of such extreme behaviours that are the reasons of extreme energy utilisation.

7 Summary and Conclusion

Increased energy demand in residential as well as in commercial buildings in recent years is deteriorating our natural energy resources and whole eco system. New and effective solutions are required to control higher rate of energy consumption in the buildings. This research work proposed a holistic multidisciplinary framework to exchange knowledge and understanding from different domains for the design and development of sustainable energy efficient buildings. This framework also presents the collaboration model to share knowledge among different stakeholders and knowledge experts to implement effective policies that help to improve energy efficiency.

This research work focuses on exploring data science techniques to understand users' energy consumption patterns in residential buildings. Electricity data is collected from 32 different residential buildings for one year. Raw data is visualized using Cumulative Distribution Function to understand its graphical distribution. However, boxplot diagrams are used to visualize outliers in the dataset. Dataset is re-sampled for different timestamp to eliminate the probability of unwanted data values. Once data was preprocessed, heatmap algorithm is designed and implement to understand electricity consumption patterns for one residential building.

Descriptive analytical method is used to elaborate the results of the heatmap. However, unusual or extreme energy utilization behavior is noticed in the energy consumption pattern and elaborated using diagnostic analytical method. Contextual analysis of the results helps to understand the rationale behind normal and unusual energy consumption patterns. Peaks were identified in the heatmap that tell us some extreme behavior of energy consumption. This, sometimes, could be because of any fault in the integrated devices at home. However, this also recommends to understand residents own behavior to use energy at home.

Energy analysis results reinforce our statement that figuring out the factors that trigger the peak energy demand for a specific period of time in a building could potentially help to improve building's heating, ventilation and air conditioning (HVAC) system. Together with this, sudden peak in energy consumption can be because of some mal-functioning or some unexceptional human behavior. Finding the possible causes of high energy demand for a certain period of time can possibly leads to find appropriate solutions for it and ultimately a control in energy demand. Understanding this demand and supply behavior in residential areas will further support the sustainable and renewable energy technology.

As part of future research work, authors intend to explore different data analytical techniques that could be used to analyze stakeholders' requirements that they want to be integrated in smart buildings.

References

1. Fan, C., Xiao, F., Wang, S.: Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. *Applied Energy*. 127, 1-10 (2014)
2. Hsu, D.: Identifying key variables and interactions in statistical models of building energy consumption using regularization. *Energy*. 83(0), 144-155 (2015).
3. Pérez-Lombard, L., Ortiz, J., and Pout, C.: A review on buildings energy consumption information. *Energy and Buildings*. 40(3), 394-398 (2008).
4. Pacala, S. and Socolow, R.: Stabilization Wedges: Solving the Climate Problem for the Next 50 Years with Current Technologies. *Science*. 305 (5686), 968-972 (2004).
5. International Energy Agency (IEA), *World Energy Outlook 2015*, OECD/IEA, Editor. 2014: Paris.
6. Kosorus, H., Honigl, J., Kung, J.: Using R, WEKA and RapidMiner in Time Series Analysis of Sensor Data for Structural Health Monitoring. In: 22nd International

Workshop on Database and Expert Systems Applications (DEXA), pp. 306–310. 29 Aug.-2 Sept., IEEE, France (2011).

7. Millan, P., et al.: Time series analysis to predict link quality of wireless community networks. *Computer Networks*. 93(2), 342-358 (2015).
8. Platchkov, L.M., Pollitt M.G.: *The Economics of Energy (and Electricity) Demand* Cambridge University. May13–14, 2011.
9. Walker, C.F., Pokoski, J.L.: Residential Load Shape Modelling Based on Customer Behavior. *IEEE Transactions on Power Apparatus and Systems*. 104(7), 1703-1711 (1985).
10. Capasso, A., et al., A bottom-up approach to residential load modeling. *IEEE Transactions on Power Systems*. 9(2), 957-964 (1994).
11. Willis, H.L. *Spatial Electric Load Forecasting*. 2nd edn, CRC Press, New York (2002)
12. Iram, S., Fernando, T., and Bassanino, M.: Exploring Cross-Domain Data Dependencies for Smart Homes to Improve Energy Efficiency. In: *Companion Proceedings of the 10th International Conference on Utility and Cloud Computing*, pp. 221-226. ACM, USA (2017).